Theoretical Computer Science 36 (1985) 109–112
North-Holland

NOTE

ON THE CHOMSKY AND STANLEY'S HOMOMORPHIC CHARACTERIZATION OF CONTEXT-FREE LANGUAGES

Sadaki HIROSE
Kanagawa University, 3-27-1, Rokkaku-bashi, Kanagawa-ku, Yokohama 221, Japan

Masaaki YONEDA
Toyama University, Takaoka 933, Japan

Communicated by A. Salomaa
Received January 1984
Revised May 1984

Abstract. In this note we refine the Chomsky and Stanley's homomorphic characterization of context-free languages: it is shown that each context-free language can be expressed in the form $h(D \cap M_R)$ for some Dyck language D, some 'minimal linear and regular' language M_R and some homomorphism h.

In formal language theory, one of the main trends is to search for characterizations for classes of languages. A typical example is the Chomsky and Stanley's homomorphic characterization of the class of context-free languages [1, 6, 3], that is, each context-free language can be expressed in the form $h(D \cap R)$ for some Dyck language D, some regular language R and some homomorphism h.

Recently, Hirose, Okawa and Yoneda proved that a result similar to the Chomsky and Stanley's theorem also holds for the class of recursively enumerable languages [4], that is, each recursively enumerable language can be expressed in the form $h(D \cap M)$, where h and D are as above and M is a minimal linear language (which is generated by a linear context-free grammar with only one nonterminal symbol [2]).

The class of regular languages is not a subclass of minimal linear languages. Hence, from the two above-mentioned theorems, the question arises whether there exists a subclass of minimal linear languages which characterizes the class of context-free languages.

In this note we give an affirmative answer of this question, that is, we show that the class of context-free languages is characterized by 'minimal linear and regular' languages.

The class of minimal linear and regular languages is a proper subclass of regular languages. Hence, our characterization of context-free languages is a refinement of Chomsky and Stanley's one.

The reader is referred to Ginsburg [3] and Salomaa [5] for background material and additional details.

First we give some definitions.

Definition 1. A grammar $G=\langle V, \Sigma, P, S\rangle$ is *minimal linear* if V contains only one nonterminal symbol (namely $V=\{S\}$) and all of the production rules in P are of the forms $S \to uSv$ and $S \to w$ for some u, v, and w in Σ^*. A language L is said to be minimal linear if $L=L(G)$ for some minimal linear grammar G.

Definition 2. A grammar $G=\langle V, \Sigma, P, S\rangle$ is *regular* if all of the production rules in P are of the forms $A \to aB$ and $A \to a$ for some A and B in V and a in $\Sigma \cup \{\varepsilon\}$. A language L is said to be regular if $L=L(G)$ for some regular grammar G.

Definition 3. A language L is said to be *minimal linear and regular* if there exist a minimal linear grammar G_1 and a regular grammar G_2 such that $L=L(G_1)=L(G_2)$.

Our main result is the following theorem.

Theorem 1. *For each alphabet Σ, there exist an alphabet Σ', a Dyck language D over Σ' and a homomorphism $h: (\Sigma')^* \to \Sigma^*$ which satisfy the property that for each context-free language L over Σ a minimal linear and regular language M_{R} over Σ' can be found such that*

$$L=h(D \cap M_{\mathrm{R}}).$$

Proof. The proof essentially differs from that of the Chomsky and Stanley's theorem (see [3]) in the construction of the grammar G' only. However, for the sake of completeness, we briefly repeat the whole construction.

Suppose that $\Sigma=\{a_1, a_2, \ldots, a_m\}$. Let

$$\Sigma'=\Sigma \cup \{a_1', a_2', \ldots, a_m'\} \cup \{c, c', d, d'\} \cup \{e, e', f, f'\}$$

be an alphabet of $2m+8$ elements. Let $G_0=\langle\{S_0\}, \Sigma', P_0, S_0\rangle$, where

$$\begin{aligned} P_0=\{&S_0 \to \varepsilon\} \\ &\cup \{S_0 \to S_0 a_i S_0 a_i' S_0 \mid 1 \leq i \leq m\} \\ &\cup \{S_0 \to S_0 c S_0 c' S_0,\ S_0 \to S_0 d S_0 d' S_0\} \\ &\cup \{S_0 \to S_0 e S_0 e' S_0,\ S_0 \to S_0 f S_0 f' S_0\}. \end{aligned}$$

Let $D=L(G_0)$. Let h be the homomorphism on $(\Sigma')^*$ defined by

$$h(a_i)=a_i, \quad h(a_i')=\varepsilon \qquad (1 \leq i \leq m),$$

$$h(c)=h(c')=h(d)=h(d')=h(e)=h(e')=h(f)=h(f')=\varepsilon.$$

Now let $L \subseteq \Sigma^*$ be an arbitrary context-free language. We may assume that $L=L(G)$ for some context-free grammar $G=\langle V, \Sigma, P, S\rangle$, where $P=$

$\{\pi_1, \pi_2, \ldots, \pi_k\}$ and each production rule π_i $(i = 1, 2, \ldots, k)$ is of the form $A \to a$ for some a in $\Sigma \cup \{\varepsilon\}$, or $A \to BC$ for some B and C in V. Let $V = \{v_1, v_2, \ldots, v_n\}$, and for v_j $(j = 1, 2, \ldots, n)$ let $C(v_j) = ef^je$ and $C'(v_j) = e'f'^je'$. Let $G' = \langle\{\sigma\}, \Sigma', P', \sigma\rangle$, where

$$\begin{aligned} P' = \{&\sigma \to C(S)\sigma\} \\ &\cup \{\sigma \to C'(A)dc^idC(B)\sigma \mid \pi_i = A \to BC,\ B \text{ and } C \text{ in } V\} \\ &\cup \{\sigma \to C'(A)aa'd'c'^id'C(D)\sigma \\ &\qquad \mid a \text{ in } \Sigma \cup \{\varepsilon\},\ A \to a \text{ in } P,\ \pi_i = B \to CD,\ C \text{ and } D \text{ in } V\} \\ &\cup \{\sigma \to C'(A)aa' \mid a \text{ in } \Sigma \cup \{\varepsilon\},\ A \to a \text{ in } P\}. \end{aligned}$$

Let $M_R = L(G')$. Since G' is minimal linear and we can easily construct a regular grammar G'_R from G' such that $M_R = L(G') = L(G'_R)$, M_R is minimal linear and regular.

As in [3], it is not difficult to show that $L = h(D \cap M_R)$. $\square$

Combining the Hirose, Okawa and Yoneda's theorem and our Theorem 1, a stronger characterization is obtained.

Theorem 2. *For each alphabet Σ, there exist an alphabet Σ', a Dyck language D over Σ' and a homomorphism $h: (\Sigma')^* \to \Sigma^*$ which satisfy the property that for each recursively enumerable language L_{RE} over Σ a minimal linear language M over Σ' can be found such that*

$$L_{RE} = h(D \cap M),$$

and for each context-free language L_{CF} over Σ a minimal linear and regular language M_R over Σ' can be found such that

$$L_{CF} = h(D \cap M_R).$$

Acknowledgment

The authors would like to thank Professor Masakazu Nasu of Mie University and Professor Satoshi Okawa of Hachinohe Institute of Technology for their valuable suggestions and discussions.

References

[1] N. Chomsky, Context-free grammars and pushdown storage, *M.I.T. Res. Lab. Electron. Quart. Prog. Report* **65** (1962) 187–194.

[2] N. Chomsky and M.P. Schützenberger, The algebraic theory of context-free languages, in: P. Braffort and D. Hirschberg, eds., *Computer Programming and Formal Systems* (North-Holland, Amsterdam, 1962) 118–161.
[3] S. Ginsburg, *The Mathematical Theory of Context-Free Languages* (McGraw-Hill, New York, 1966).
[4] S. Hirose, S. Okawa and M. Yoneda, A homomorphic characterization of recursively enumerable languages, *Theoret. Comput. Sci.* **35** (2, 3) (1985) 261–269.
[5] A. Salomaa, *Formal Languages* (Academic Press, New York, 1973).
[6] R.J. Stanley, Finite state representations of context-free languages, *M.I.T. Res. Lab. Electron. Quart. Prog. Report* **76** (1965) 276–279.